IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jefferson Valandro
20/12/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies

- **Data Collection:** Comprehensive gathering of information about SpaceX, including historical launch data, mission parameters, and any publicly available details on first stage landings.
- **Dashboard Creation:** Utilization of gathered data to construct informative dashboards. These dashboards serve as visual tools for the Space Y team, offering insights into SpaceX's launch costs and first stage reusability.
- **Machine Learning Training:** A machine learning model is trained using the acquired data to predict the likelihood of successful first stage landings. Publicly available information becomes the basis for training the model to enhance accuracy.

- Summary of all results

- **Launch Cost Determination:** The project successfully determines the cost of SpaceX launches by considering various factors, including first stage reusability. Insights gained from historical data contribute to a comprehensive understanding of pricing dynamics .
- **First Stage Reusability Prediction:** Through the implementation of machine learning, the project achieves a predictive model capable of forecasting the probability of the Falcon 9 first stage being reused. This information aids Space Y in strategic decision -making and competition with SpaceX.

Introduction

Project Background and Context

This project delves into the dynamic realm of commercial space travel, where companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX have revolutionized accessibility to space exploration. Notably, SpaceX stands out for its remarkable achievements in sending missions to the International Space Station, deploying the Starlink satellite internet constellation, and conducting manned space missions.

The cost-effectiveness of SpaceX's Falcon 9 rocket launches, priced at $62 million compared to other providers' $165 million cost per launch, primarily arises from their innovative first stage reusability. This project aims to determine the cost of a launch by predicting the successful landing and subsequent reuse of the Falcon 9's first stage.

Problems Seeking Solutions

- Cost Prediction: Determining the cost of a launch by forecasting the first stage's landing, crucial in estimating expenses.

- First Stage Reusability: Training a machine learning model using public information to predict if SpaceX will reuse the first stage, thereby enhancing insights into cost assessments.

Section 1

Methodology

Methodology

Executive Summary

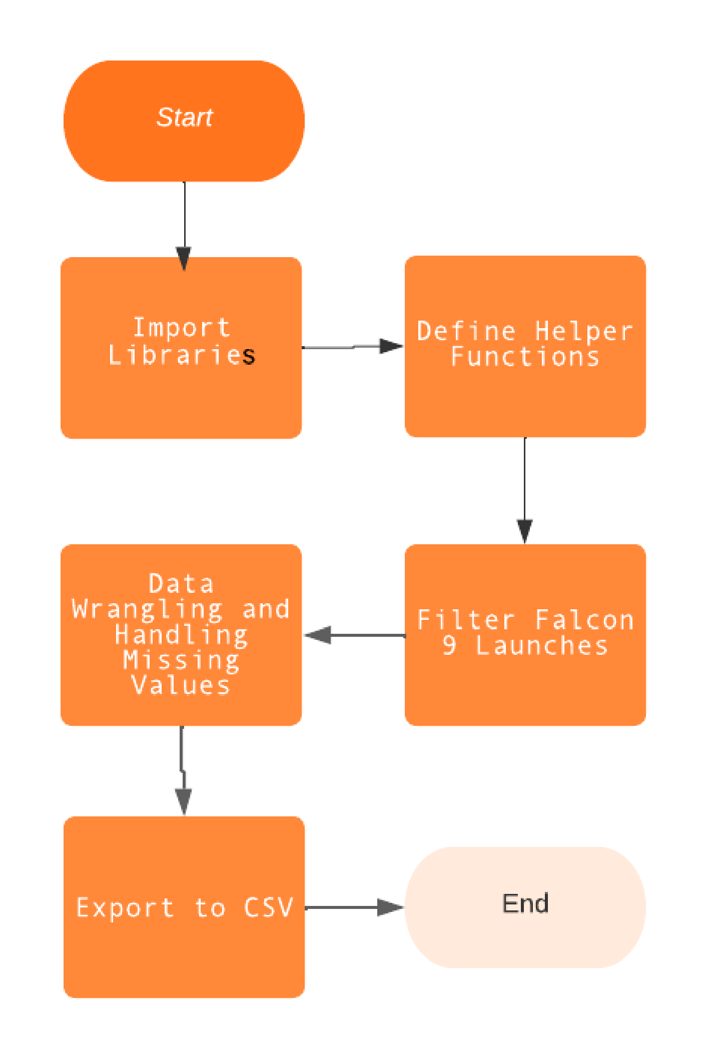
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

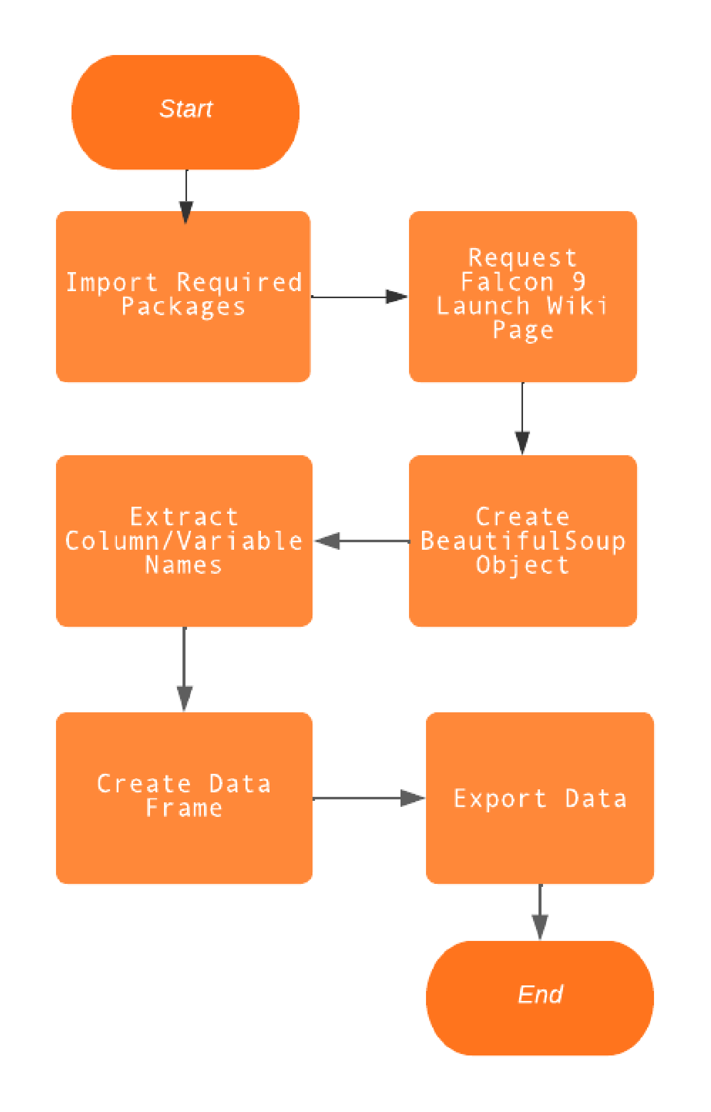
- **Import Libraries and Define Auxiliary Functions:**
  - Imported necessary libraries like requests, pandas, numpy, and datetime.
  - Defined helper functions for extracting information from the SpaceX API based on rocket, launchpad, payload, and cores.
- **Request and Parse SpaceX Launch Data:**
  - Made a GET request to the SpaceX API to retrieve past launch data.
  - Used static JSON data for consistency in responses.
  - Checked the success of the request with a 200 status code.
  - Decoded the response content as JSON and converted it into a Pandas dataframe using json_normalize.
  - Displayed the first 5 rows of the dataframe.
- **Filter Falcon 9 Launches:**
  - Filtered the dataframe to include only Falcon 9 launches based on the BoosterVersion column.
  - Reset the FlightNumber column to start from 1.
- **Data Wrangling and Handling Missing Values:**
  - Handled missing values in the PayloadMass column by replacing them with the mean.
  - Identified and acknowledged missing values in the LandingPad column.
- **Export to CSV:**
  - Exported the filtered Falcon 9 dataset to a CSV file named 'dataset_part_1.csv'.
- Here you can check the notebook.



Data Collection - Scraping

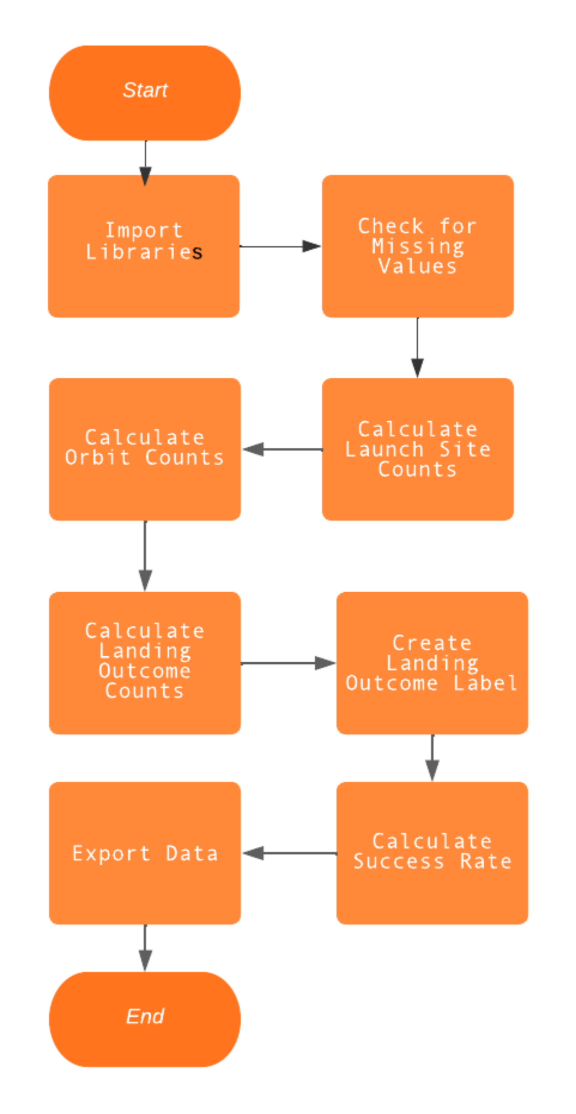
- **Import Required Packages:**
  - Import necessary Python packages for web scraping.
- **Request Falcon 9 Launch Wiki Page:**
  - Use `requests.get()` to fetch the Falcon 9 Launch HTML page.
  - Assign the response to an object.
- **Create BeautifulSoup Object:**
  - Use BeautifulSoup to create an object from the HTML response.
  - Print the page title to verify.
- **Extract Column/Variable Names:**
  - Find tables on the page and extract column names from the third table.
- **Create Data Frame:**
  - Create an empty dictionary with keys from the column names.
  - Initialize the dictionary with empty lists.
  - Fill up the dictionary with launch records extracted from table rows.
- **Export Data:**
  - Save the DataFrame to a CSV file for future use.

- You can see the notebook here.



Data Wrangling

- **Import Libraries and Load Data:**
  - Utilize Pandas and NumPy libraries for data manipulation and analysis.
  - Load the SpaceX dataset from the provided URL.
- **Check for Missing Values:**
  - Identify and calculate the percentage of missing values in each attribute.
- **Identify Data Types:**
  - Determine which columns are numerical and categorical.
- **Calculate Launch Site Counts:**
  - Count the number of launches for each launch site.
- **Calculate Orbit Counts:**
  - Count the number of launches for each orbit type.
- **Calculate Landing Outcome Counts:**
  - Count the number of landing outcomes.
- **Create Landing Outcome Label:**
  - Create a binary label for landing outcomes.
- **Calculate Success Rate:**
  - Calculate the success rate of the first stage landing.
- **Export Data:**
  - Save the processed data to a new CSV file.
- You can access the notebook [here].



EDA with Data Visualization

- **FlightNumber vs. PayloadMass:**
  - Chart Type: Catplot
  - Reason: To visualize the relationship between the continuous launch attempts (FlightNumber) and the payload mass, overlaying the launch outcome. It helps observe trends and patterns as the flight number increases.
- **Launch Site Success Rates:**
  - Chart Type: Barplot
  - Reason: To compare success rates at different launch sites. It provides a clear visual representation of the success rates for each launch site.
- **FlightNumber vs. Launch Site:**
  - Chart Type: Scatter Plot
  - Reason: To explore the relationship between the flight number and launch site. It helps identify any patterns or trends in the distribution of launches across different flight numbers and launch sites.
- **Payload vs. Launch Site:**
  - Chart Type: Scatter Plot
  - Reason: To observe if there is any relationship between launch sites and their payload mass. It helps identify any patterns or correlations between payload mass and launch sites.
- **Success Rate by Orbit Type:**
  - Chart Type: Barplot
  - Reason: To visualize the success rate for each orbit type. It provides insights into the success rates associated with different types of orbits.
- **FlightNumber vs. Orbit Type:**
  - Chart Type: Scatter Plot
  - Reason: To explore the relationship between the flight number and orbit type. It helps identify patterns or trends in the distribution of launches across different flight numbers and orbit types.
- **Payload vs. Orbit Type:**
  - Chart Type: Scatter Plot
  - Reason: To observe the relationship between payload mass and orbit type. It helps identify any patterns or correlations between payload mass and different orbit types.
- **Yearly Launch Success Trend:**
  - Chart Type: Barplot and Line Plot
  - Reason: To visualize the yearly trend in launch success rates. Barplot shows the success rate for each year, and the line plot shows the average yearly success trend.
- Here you can find the notebook

EDA with SQL

- **Task 1: Display Unique Launch Sites**
  - Displayed the names of unique launch sites.
- **Task 2: Display Records with Launch Sites Starting with 'CCA'**
  - Displayed 5 records where launch sites begin with 'CCA'.
- **Task 3: Display Total Payload Mass for NASA (CRS) Launches**
  - Displayed the total payload mass carried by boosters launched by NASA (CRS).
- **Task 4: Display Average Payload Mass for Booster Version F9 v1.1**
  - Displayed the average payload mass carried by booster version F9 v1.1.
- **Task 5: Display Date of First Successful Landing Outcome on Ground Pad**
  - Displayed the date when the first successful landing outcome on the ground pad was achieved.
- **Task 6: Display Booster Versions with Success in Drone Ship and Payload Mass Range**
  - Displayed the names of booster versions with success in the drone ship and payload mass between 4000 and 6000.
- **Task 7: Display Total Number of Successful and Failure Mission Outcomes**
  - Displayed the total number of successful and failure mission outcomes.
- **Task 8: Display Booster Versions with Maximum Payload Mass**
  - Displayed the names of booster versions that carried the maximum payload mass.
- **Task 9: Display Records for Months in 2015 with Failure Landing Outcomes in Drone Ship**
  - Displayed records for months in 2015 with failure landing outcomes in the drone ship.
- **Task 10: Rank Landing Outcomes Count in Descending Order for a Date Range**
  - Ranked the count of landing outcomes in descending order for a specified date range.
- Here you can find the notebook.

Build an Interactive Map with Folium

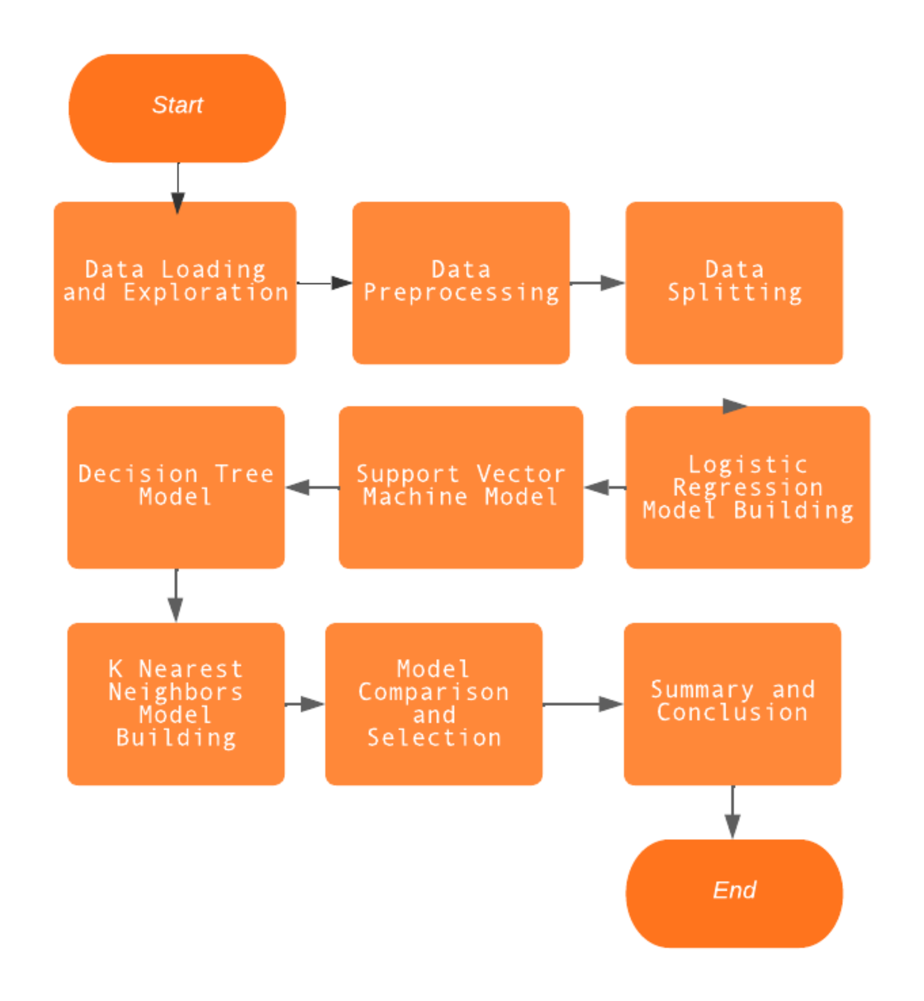
- **Markers:**
  - Created markers for each launch site with its coordinates (latitude and longitude).
  - Used different marker colors (green for success, red for failure) to indicate launch outcomes.
  - Added these markers to a MarkerCluster to handle multiple markers at the same location efficiently.
- **Circles:**
  - Added circles to represent the launch sites, highlighting their general areas on the map.
  - Used circles to provide a visual representation of the proximity of launch sites to certain locations.
- **PolyLine:**
  - Drew polylines between launch sites and selected proximity points (coastline in the example) to visualize the distances.
- **MousePosition:**
  - Added a MousePosition plugin to get the latitude and longitude coordinates when hovering over the map.
  - Enabled easy exploration and identification of specific points on the map.
- **Distance Markers:**
  - Created markers to indicate the distance between launch sites and specific points like coastlines.
  - Used these markers to display the calculated distances as labels on the map.
  - **Explanation:**
- **Markers:** Used to mark the exact locations of launch sites, making it easy to identify them on the map. The use of different colors helps distinguish successful and failed launches.
- **Circles:** Provided a visual representation of the launch site areas, making it easier to observe their geographical distribution.
- **PolyLine:** Used to draw lines between launch sites and proximity points, visually representing the distances and connections.
- **MousePosition:** Enabled users to explore the map and find precise coordinates for specific points of interest.
- **Distance Markers:** Displayed the calculated distances between launch sites and selected proximity points, providing additional information about their spatial relationships.
- Here you can find the notebook.

Build a Dashboard with Plotly Dash

- **Launch Site Dropdown:**
  - Interacts with pie charts and scatter plots.
  - Allows selection of specific launch sites or all sites for analysis.
- **Success Pie Chart:**
  - Displays the success count for each launch site or overall.
  - Provides a clear visual comparison of successful launches per site or in total.
  - Helps identify the site with the largest successful launches and the highest launch success rate.
- **Payload Range Slider:**
  - Interacts with the scatter plot.
  - Enables the selection of payload ranges for observation.
  - Facilitates the examination of launch success rates concerning payload mass.
- **Success Payload Scatter Chart:**
  - Represents the relationship between payload mass and launch success.
  - Displays the scatter points for payload mass against launch outcomes (success/failure) for selected sites or all sites.
  - Color-labels the points based on the F9 Booster version.
  - Aids in understanding correlations between payload mass, launch outcomes, and booster versions.
  - **Reasons for Adding Plots and Interactions:**
- **Dropdown and Pie Chart:**
  - Allow quick comparison of success counts and rates between different launch sites.
  - Enable identification of the site with the largest successful launches and the highest success rate, providing insights into site performance.
- **Payload Range Slider and Scatter Chart:**
  - Facilitate the analysis of the relationship between payload mass and launch success.
  - Help identify payload ranges with the highest and lowest launch success rates.
  - Color-labelling based on booster versions allows assessment of how different versions impact launch success.
  - These plots and interactions were incorporated to provide a comprehensive analysis of SpaceX launch data. The visualizations help in identifying patterns, trends, and correlations, offering insights into factors influencing launch success and site performance.

Predictive Analysis (Classification)

- **Data Loading and Exploration:**
  - Loaded dataset from provided URLs using Pandas.
  - Explored the data structure, features, and target variable.
- **Data Preprocessing:**
  - Created a Pandas series 'Y' by extracting the 'Class' column as the target variable.
  - Standardized the feature data in 'X' using the StandardScaler from sklearn.
- **Data Splitting:**
  - Split the data into training and test sets using the train_test_split function.
  - Used 80% of the data for training and 20% for testing.
- **Logistic Regression Model:**
  - Created a logistic regression object.
  - Utilized GridSearchCV to find the best hyperparameters for logistic regression.
  - Evaluated accuracy on the test data and displayed the confusion matrix.
- **Support Vector Machine Model:**
  - Created a support vector machine object.
  - Used GridSearchCV to find the best hyperparameters for SVM.
  - Evaluated accuracy on the test data and displayed the confusion matrix.
- **Decision Tree Model:**
  - Created a decision tree classifier object.
  - Employed GridSearchCV to find the best hyperparameters for the decision tree.
  - Evaluated accuracy on the test data and displayed the confusion matrix.
- **K Nearest Neighbors Model:**
  - Created a k-nearest neighbors object.
  - Utilized GridSearchCV to find the best hyperparameters for KNN.
  - Evaluated accuracy on the test data and displayed the confusion matrix.
- **Model Comparison and Selection:**
  - Compared the performance of logistic regression, SVM, decision tree, and KNN.
  - Considered accuracy scores and confusion matrices for each model.
  - Selected the model with the highest accuracy on the test data.
- **Summary and Conclusion:**
  - Summarized the entire process, including data preparation, model training, hyperparameter tuning, and evaluation.
  - Provided insights into the best-performing classification model based on accuracy and confusion matrix analysis.
- Here you can find the notebook.



Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

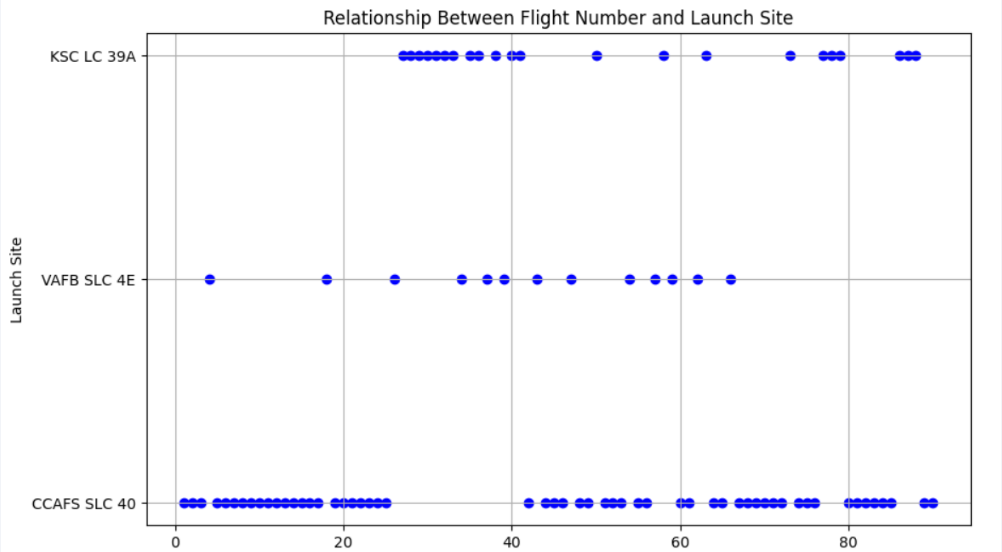
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

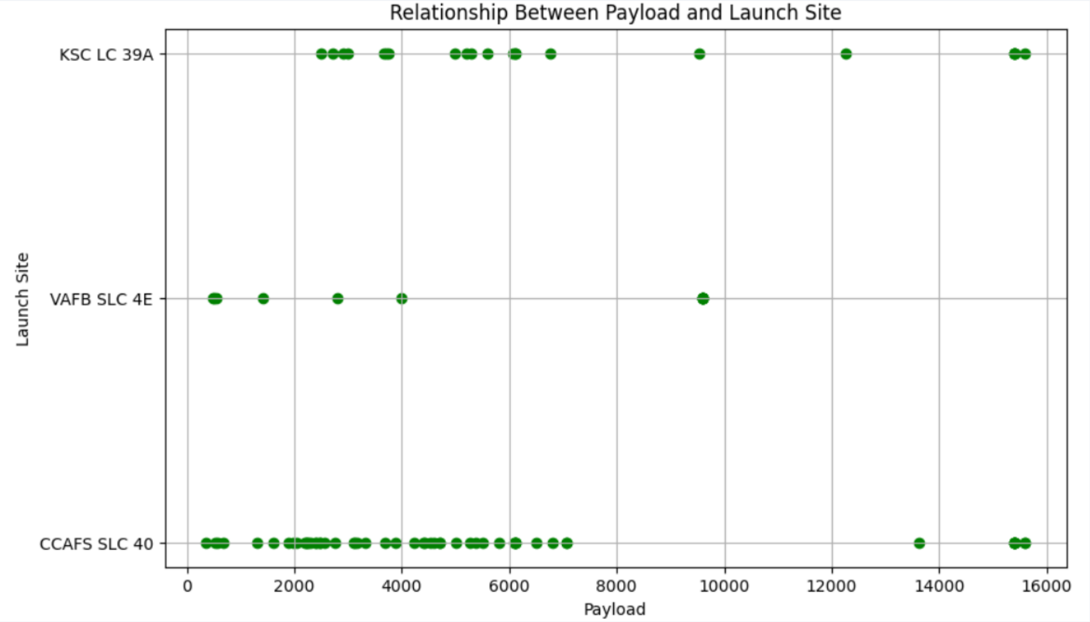
- **Distribution of Launches Across Sites:**
- The scatter plot visually represents the distribution of SpaceX launches across different launch sites.
- Each point on the plot corresponds to a specific flight, and the x-axis shows the Flight Number.
- **Flight Number Trend:**
- Observe the trend of Flight Numbers over the launches. Does the Flight Number increase consistently, or are there variations?
- The x-axis (Flight Number) helps in understanding the chronological order of launches.

- **Analyzing the scatter plot of Flight Number versus Launch Site reveals a notable concentration of launches at one specific site, indicating significantly more launches compared to the other two. Additionally, there is a distinct difference in the initiation of launch activities, with one site showing early and consistent launches, while the others exhibit a more gradual growth curve. This disparity underscores the need to understand factors influencing launch success and highlights opportunities for strategic adjustments in operations to optimize performance based on evolving space industry demands.**



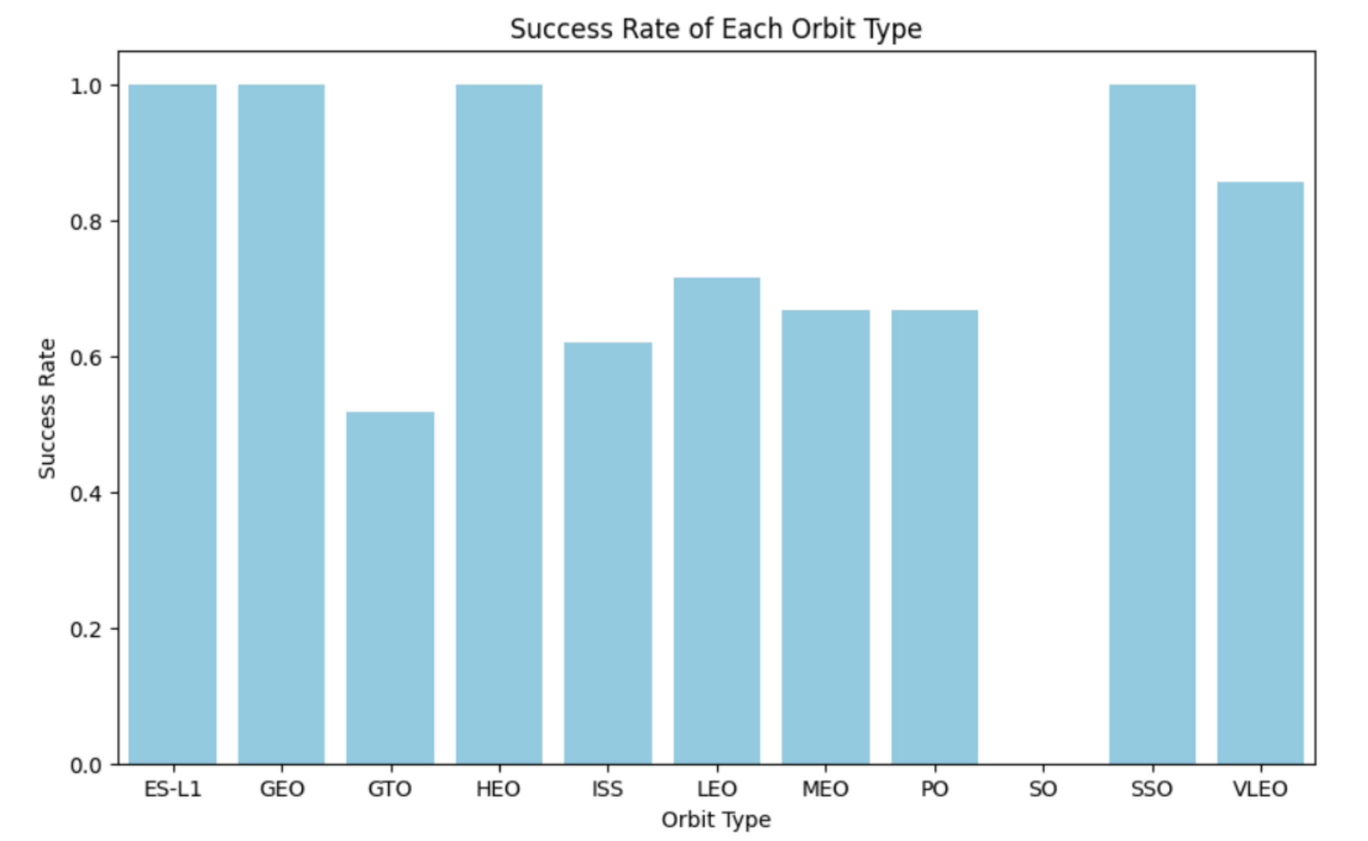
Payload vs. Launch Site

- **Payload Distribution Across Launch Sites:**
  - The scatter plot visually represents the distribution of payload masses for SpaceX launches across different launch sites.
- **Launch Site Comparison:**
  - For the selected launch site or when considering all sites, the plot allows us to compare how payload masses vary. This comparison provides insights into the capacity and preferences of each launch site.
- **Operational Insights:**
- Over time, patterns in the scatter plot may reveal operational adjustments, such as an increase in the payload capacity of a particular launch site or the introduction of new rocket variants.
- **Decision Support for Payload Selection:**
  - Companies or organizations planning satellite launches can use this plot to make informed decisions about which launch site aligns best with their payload requirements.
- **Temporal Trends:**
  - Analyzing the scatter plot over different time periods may reveal temporal trends in payload masses, indicating advancements in technology or changes in mission objectives.



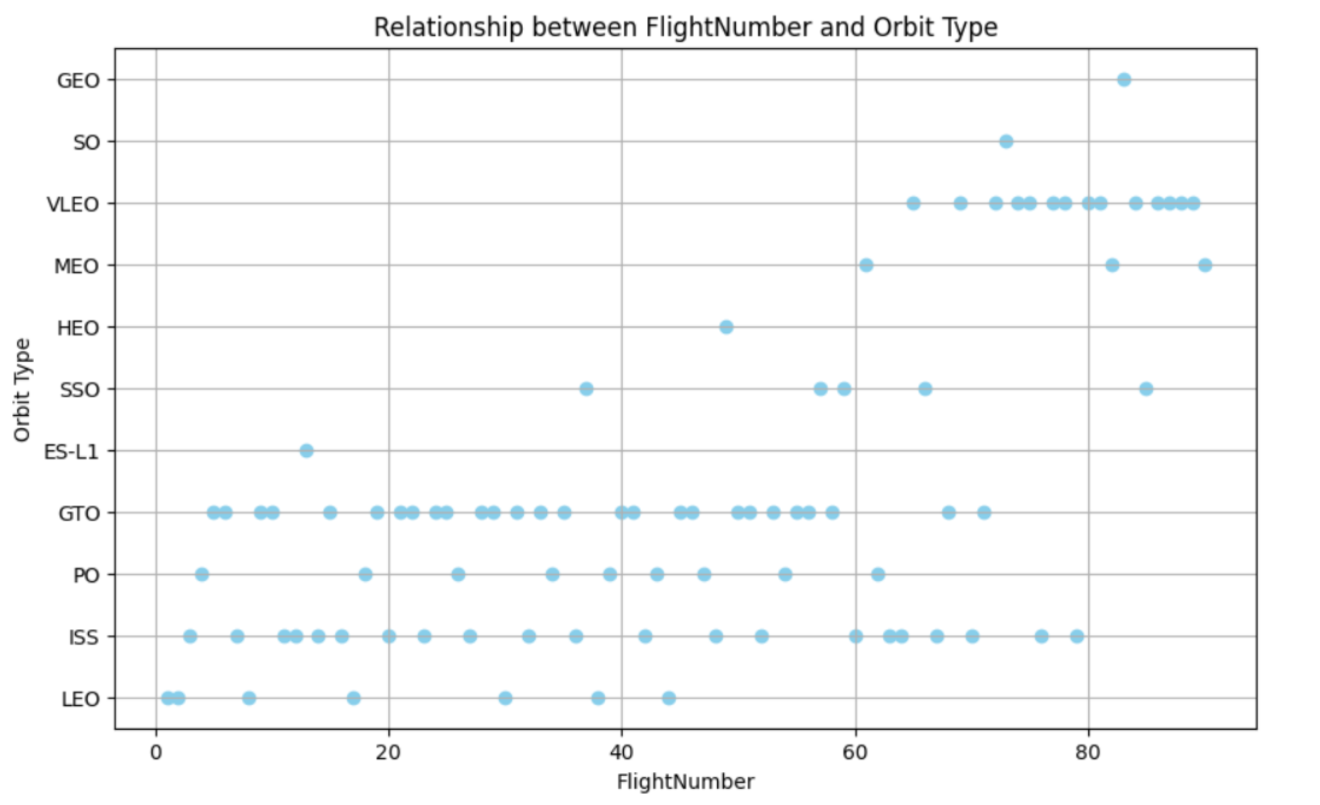
Success Rate vs. Orbit Type

- This analysis enables a comprehensive understanding of how success rates vary across distinct orbit types, aiding in strategic decision-making for future space missions.
  The bar chart illustrates the success rates across different orbit types. It's evident that some orbits exhibit significantly higher success rates compared to others. For instance, (GEO, ES-L1, HEO, SSO) demonstrate remarkable success rates, indicating a higher reliability or favorable conditions for launches. Conversely, (SO, GTO, ISS), in contrast, show a complete absence of successful launches, suggesting potential challenges or complexities associated with those orbits. This visualization provides critical insights into the success distribution across various orbit types, highlighting the disparities and indicating areas that might require further investigation or improvement for successful missions.



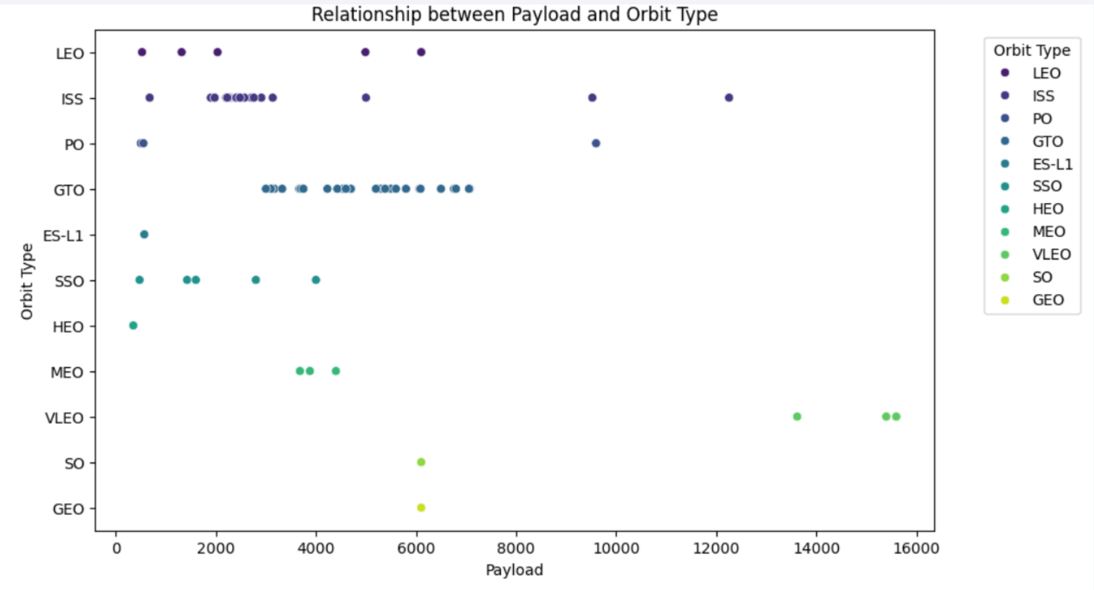
Flight Number vs. Orbit Type

- **Consecutive Launches for Specific Orbits:** Certain orbit types may witness consecutive launches, forming streaks or clusters along the Flight Number axis. This could imply a focused series of missions targeting a particular orbit type.
- **Sparse Launches for Certain Orbits:** Some orbit types may have sparse data points, suggesting fewer launches or intermittent missions targeting those specific orbits.
- **Evolution of Launch Programs:** The plot might reveal the evolution of space programs over time. For example, initial flights could be concentrated in specific orbit types, with diversification occurring in later flights.
- Overall, the Flight Number versus Orbit Type scatter plot serves as a visual representation of the historical distribution of launches across different orbits, offering valuable insights into the trends and patterns of space missions.



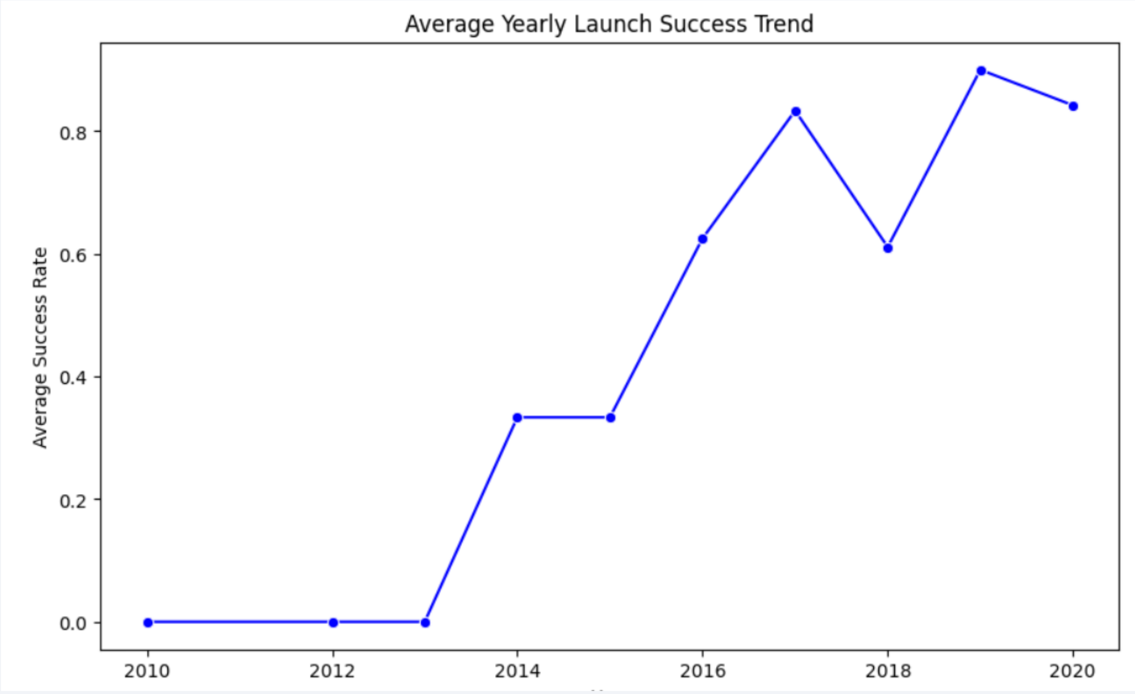
Payload vs. Orbit Type

- **High Success Rate Orbits:** Some orbit types exhibit significantly higher success rates, indicating a track record of successful launches. This could be attributed to various factors such as the complexity of the orbit, mission planning, or the reliability of the launch vehicle.
- **Varied Success Rates:** Disparities in bar heights highlight variations in success rates among different orbit types. Certain orbits may have a consistent record of success, while others may experience a mix of successful and unsuccessful launches.
- **Orbits with No Success:** It's noteworthy if there are orbit types with zero success rates. This could be due to the inherent challenges associated with those orbits or a lack of missions targeting those specific trajectories.
- **Insights into Mission Planning:** The chart provides insights into the strategic planning of space missions. Mission planners may prioritize certain orbits based on their historical success rates, ensuring a higher likelihood of mission success.
- **Identifying Risky Orbits:** Orbits with lower success rates may indicate challenges or complexities, making them riskier for space missions. Understanding these patterns is crucial for decision-making in future mission planning.
- In summary, the bar chart effectively communicates the success rates associated with different orbit types, offering valuable information for space agencies and mission planners to optimize their strategies and enhance mission success probabilities.



Launch Success Yearly Trend

- **Technological Advancements:** The increasing trend may be attributed to advancements in space technology, including improvements in launch vehicle design, navigation systems, and safety measures. This suggests that the space agency has been successful in adopting and implementing cutting-edge technologies.
- **Experience and Expertise:** Over the years, space agencies accumulate valuable experience and expertise in mission planning, execution, and problem-solving. The increasing success rate could reflect the growing proficiency of the teams involved in space missions.
- **Quality Assurance Practices:** Enhanced quality assurance practices and rigorous testing procedures contribute to a higher success rate. The space agency may have implemented stricter quality control measures to ensure the reliability of launch vehicles and components.
- **Learning from Failures:** Failures in previous missions often lead to thorough investigations and learning opportunities. The space agency may have effectively addressed and learned from past failures, implementing corrective actions to prevent similar issues in subsequent missions.
- **Improved Risk Management:** A rising success rate suggests improved risk management strategies. The space agency may have become more adept at identifying and mitigating potential risks, resulting in a more reliable and successful space program.
- **International Collaboration:** Collaborations with international space agencies and organizations can contribute to shared knowledge and best practices. The increasing success rate may be influenced by collaborative efforts and the exchange of expertise within the global space community.
- **Strategic Planning:** Strategic planning and mission prioritization play a crucial role in mission success. The space agency may have refined its strategic planning processes, focusing on missions with higher probabilities of success.



All Launch Site Names

- Find the names of the unique launch sites

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

In [14]:
```
%%sql
SELECT * FROM SPACEXTABLE
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[14]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Display the total payload mass carried by boosters launched by NASA (CRS)

In [17]:
```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") AS TotalPayloadMass
FROM SPACEXTABLE
WHERE "Customer" = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[17]:

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Display average payload mass carried by booster version F9 v1.1

In [18]:
```
%%sql

SELECT AVG("PAYLOAD_MASS__KG_") AS AveragePayloadMass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[18]:

AveragePayloadMass
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

In [19]:

```
%%sql

SELECT MIN("Date") AS FirstSuccessfulLandingDate
FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE 'Success%Ground%';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[19]:

FirstSuccessfulLandingDate
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [20]:

```
%%sql

SELECT "Booster_Version"
FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE 'Success%Drone%'
  AND "PAYLOAD_MASS__KG_" > 4000
  AND "PAYLOAD_MASS__KG_" < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[20]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

In [22]:
```
%%sql

SELECT "Mission_Outcome", COUNT(*) AS TotalCount
FROM SPACEXTABLE
GROUP BY "Mission_Outcome";
```

```
 * sqlite:///my_data1.db
Done.
```

Out[22]:

Mission_Outcome	TotalCount
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

In [24]:

```
%%sql

SELECT "Booster_Version"
FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (
    SELECT MAX("PAYLOAD_MASS__KG_")
    FROM SPACEXTABLE
);
```

```
 * sqlite:///my_data1.db
Done.
```

Out[24]:

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [27]:

```
%%sql

SELECT
    CASE strftime('%m', "Date")
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date", 0, 5) = '2015'
    AND "Landing_Outcome" LIKE 'Failure%Drone%';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[27]:

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

In [28]:

```
%%sql

SELECT
    "Landing_Outcome",
    COUNT(*) AS OutcomeCount
FROM SPACEXTABLE
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY OutcomeCount DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

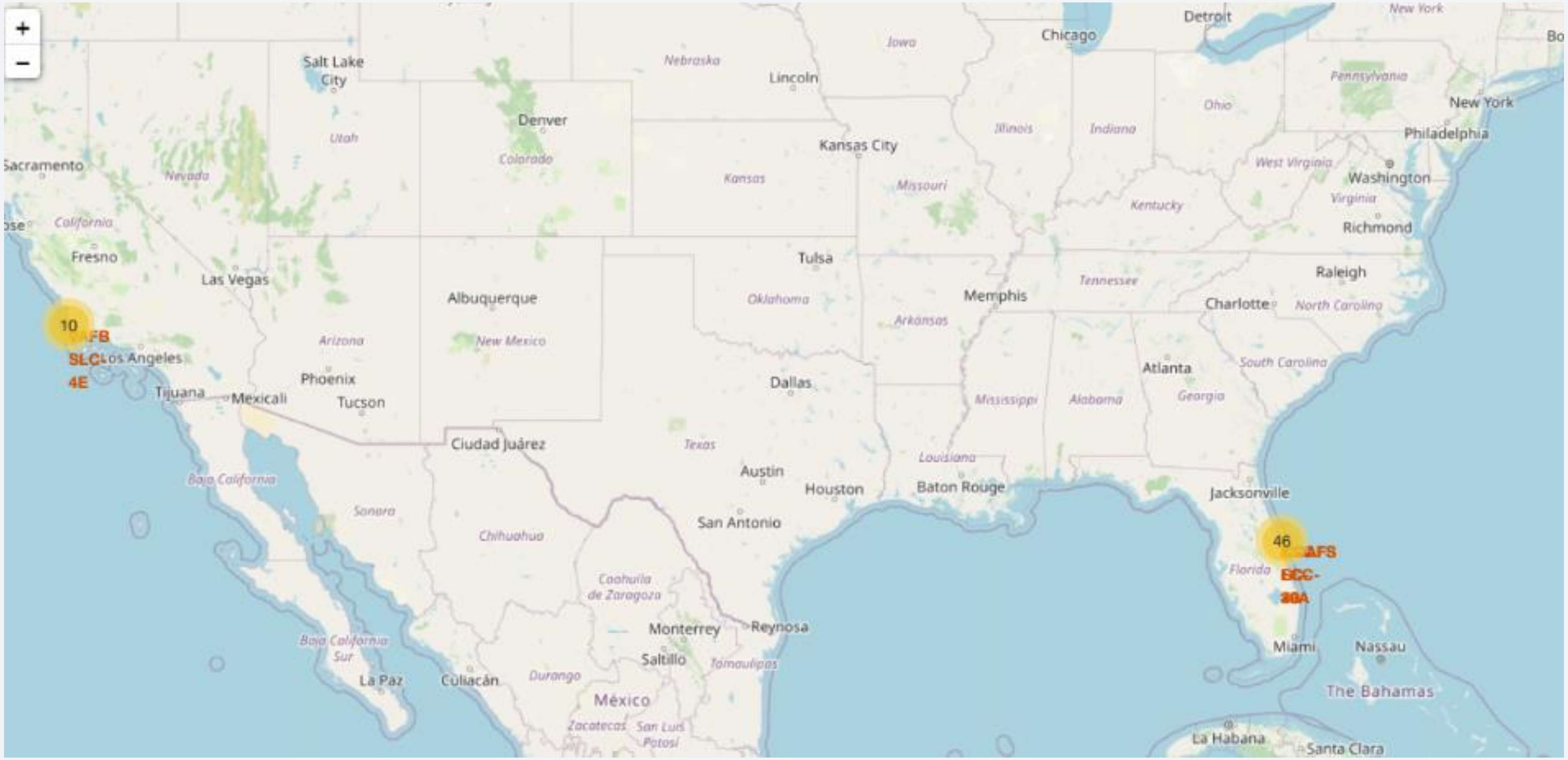
Out[28]:

Landing_Outcome	OutcomeCount
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

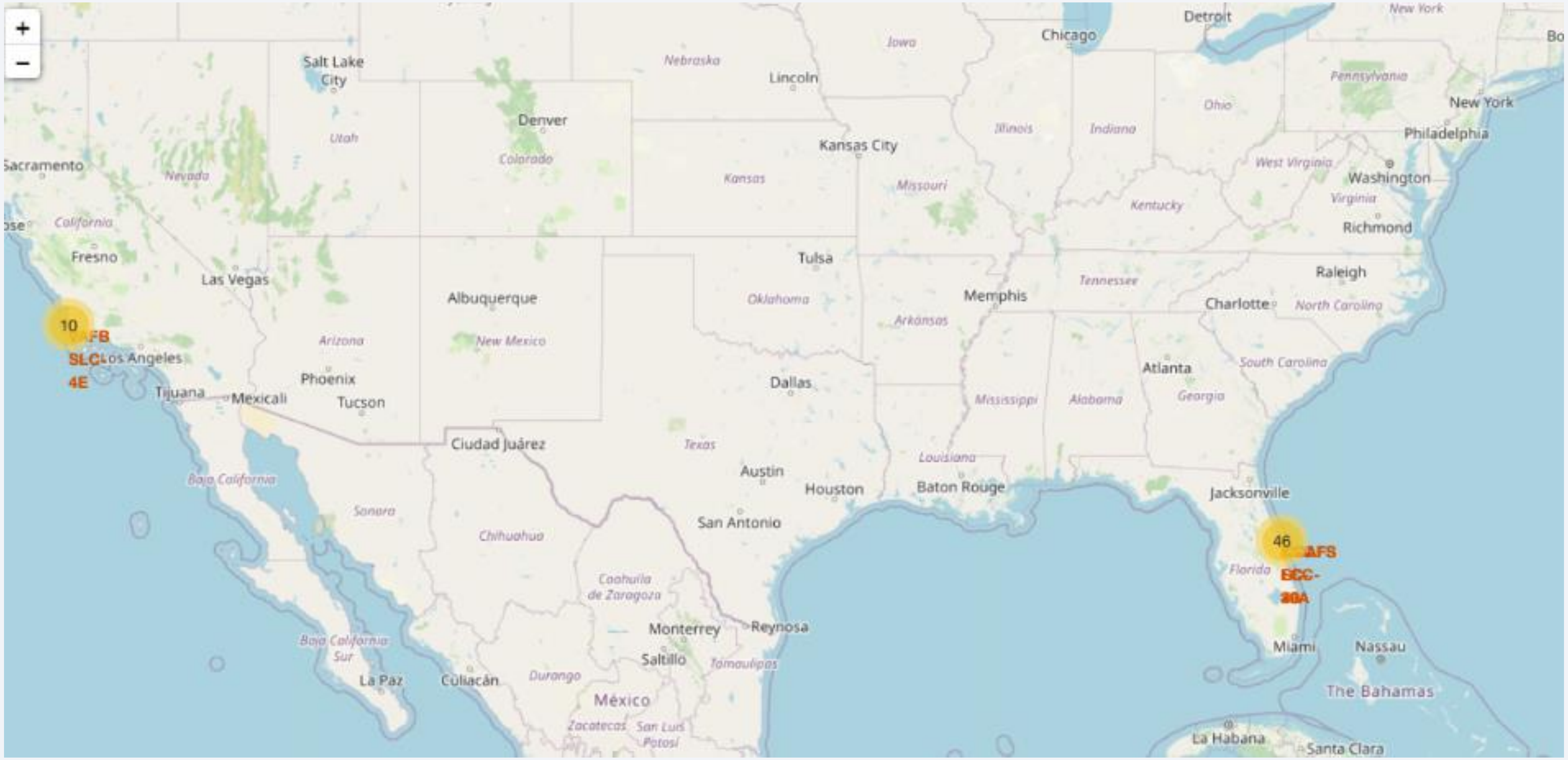
Section 3

Launch Sites Proximities Analysis

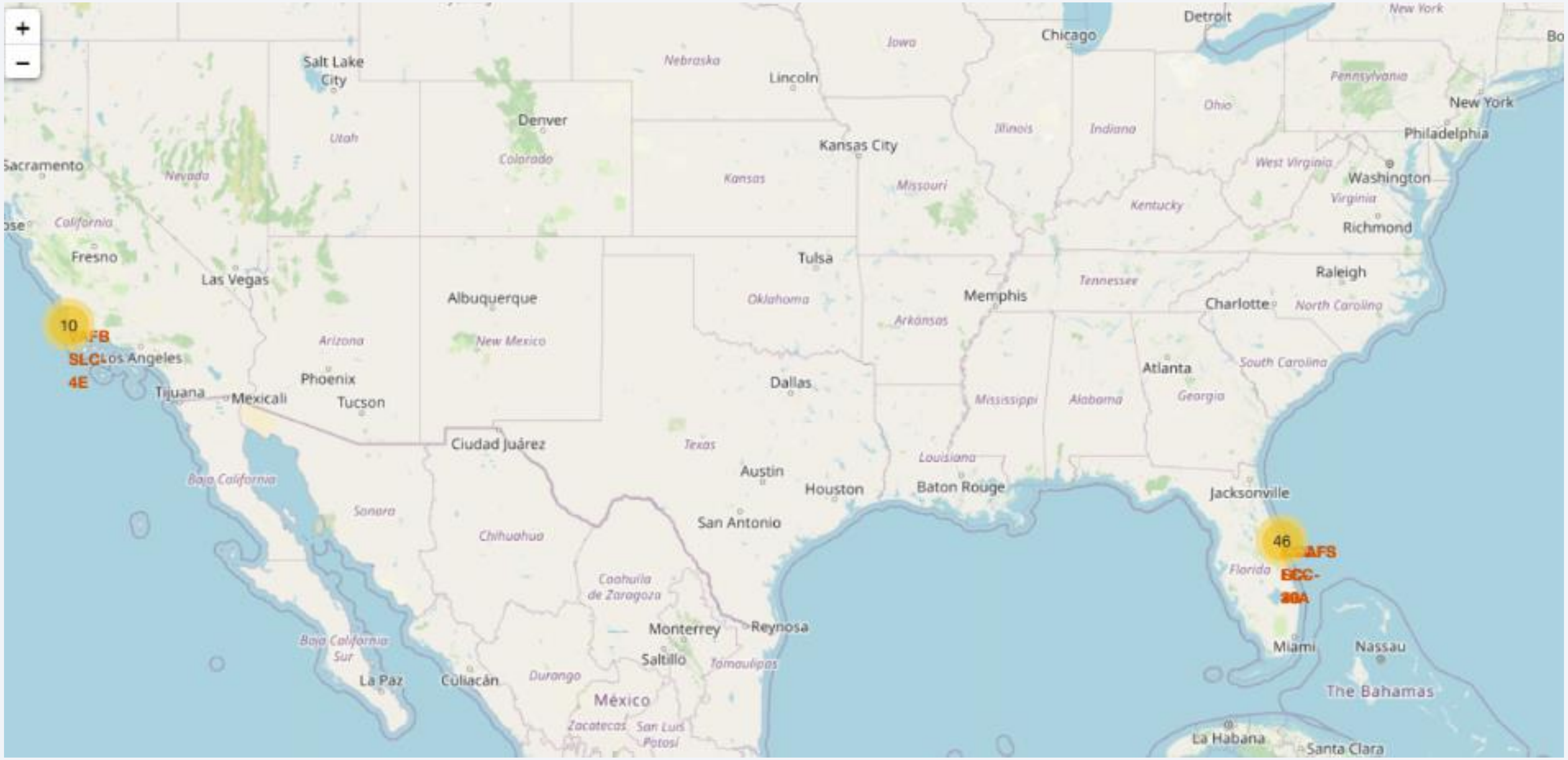
Global Launch Sites Map: Exploring Space Mission Locations



Global Launch Sites Map: Exploring Space Mission Locations



Global Launch Sites Map: Exploring Space Mission Locations



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title

- Show the screenshot of launch success count for all sites, in a piechart

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

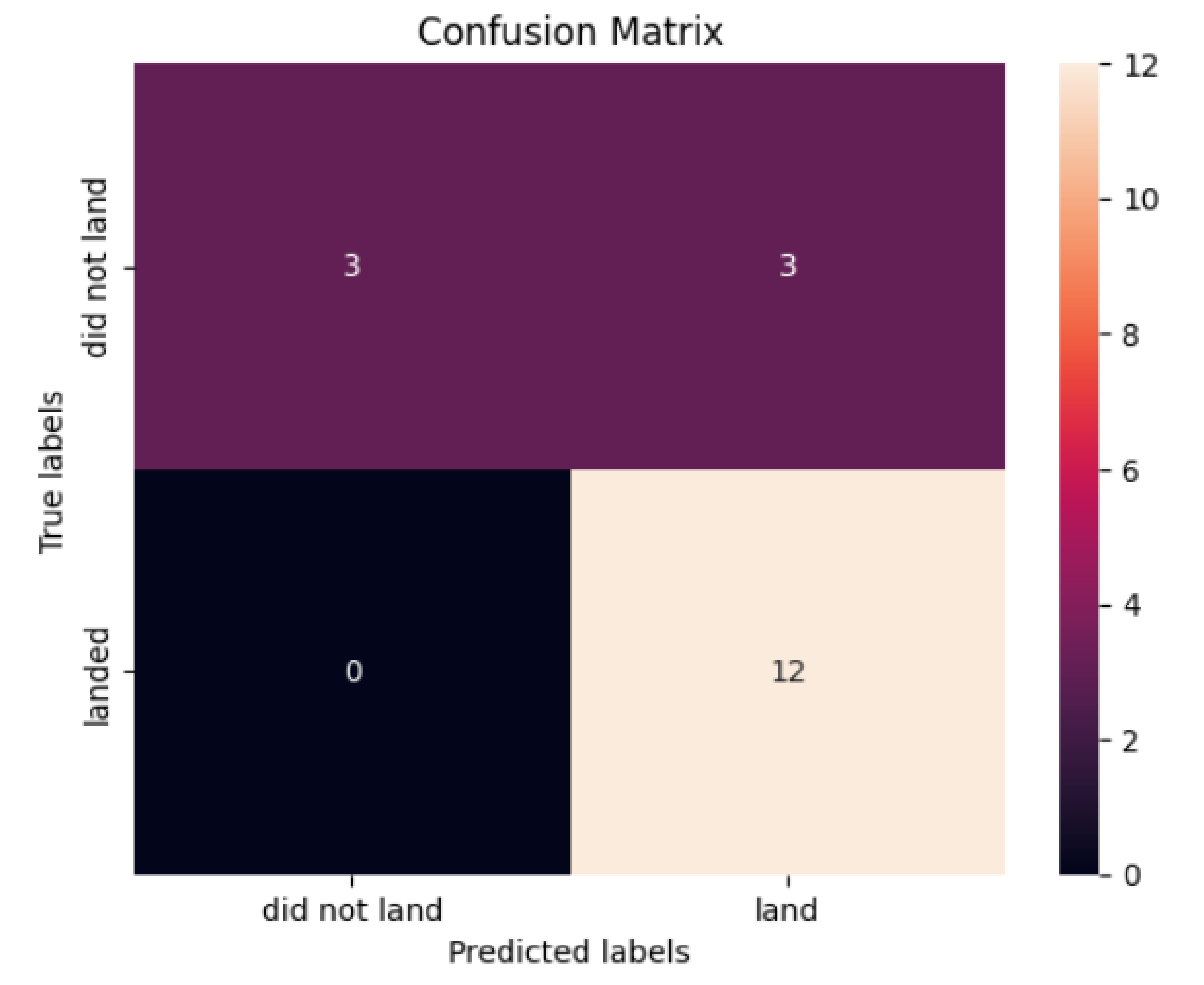
Predictive Analysis (Classification)

Classification Accuracy

```
Accuracy on Test Data:
Logistic Regression: 0.8333333333333334
Support Vector Machine: 0.8333333333333334
Decision Tree: 0.6666666666666666
K Nearest Neighbors: 0.8333333333333334

The best-performing model is: Logistic Regression with accuracy: 0.8333333333333334
```

Confusion Matrix



Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Conclusion

```
Accuracy on Test Data:
Logistic Regression: 0.8333333333333334
Support Vector Machine: 0.8333333333333334
Decision Tree: 0.6666666666666666
K Nearest Neighbors: 0.8333333333333334

The best-performing model is: Logistic Regression with accuracy: 0.8333333333333334
```

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!